2. (Amended) A method as recited in Claim 1, further comprising the steps of:
modifying a second acoustic model of the second phoneme by moving at least one mean value thereof farther from the feature values used to score the second phoneme.

A1

3. (Amended) A method as recited in Claim 1, wherein receiving correct alignment data comprises the step of receiving correct alignment data that represents a segment alignment of a less than highest scoring hypothesis from among n-best hypotheses of an utterance that was received by the speech recognition system.

---

4. (Unamended) A method as recited in Claim 1, wherein receiving wrong alignment data comprises the steps of receiving wrong alignment data that represents an alignment of the utterance that is known to be incorrect based on user confirmation information received from the speech recognition system in response to prompting a speaker to confirm the utterance.

5. (Unamended) A method as recited in Claim 1, wherein receiving correct alignment data comprises the steps of receiving correct alignment data that represents an alignment of the utterance that is known to be correct based on user confirmation information received from the speech recognition system in response to prompting a speaker to confirm the utterance.

---

6. (Amended) A method as recited in Claim 1, further comprising the step of iteratively repeating the identifying and modifying steps for all phonemes in the correct alignment data that correspond to one or more phonemes in the wrong alignment data.

B2

7. (Amended) A method as recited in Claim 2, further comprising the step of iteratively repeating the identifying and modifying steps for all phonemes in the wrong alignment data that correspond to one or more phonemes in the correct alignment data.

8. (Amended) A method as recited in Claim 1, wherein the step of moving at least one mean value farther from the feature value used to score it comprises subtracting a multiple of the feature value from the mean value of the first acoustic model.

9. (Amended) A method as recited in Claim 1, wherein the step of moving at least one mean value farther from the feature value used to score it comprises modifying the mean value of the first acoustic model by approximately two percent (2%).

10. (Amended) A method as recited in Claim 1, wherein the first acoustic model includes a plurality of model components and wherein modifying a first acoustic model further comprises the steps of modifying a set of the model components associated with the first phoneme by moving all mean values thereof closer to the corresponding feature values used to score the phoneme.

11. (Amended) A method as recited in Claim 2, wherein the second acoustic model includes a plurality of model components and wherein modifying the second acoustic model further comprises the steps of modifying a set of model components associated with the second phoneme by moving all mean values thereof farther from the corresponding feature values used to score the phoneme.

12. (Amended) A method of improving performance of a segmentation-based automatic speech recognition system (ASR) by training its acoustic models using information obtained from a particular application in which the ASR is used, comprising the steps of:

receiving a correct segment alignment of an utterance that was received by the ASR;

receiving an incorrect alignment of the utterance that is known to be incorrect based on information received from the speech recognition system in the context of the particular application;

identifying a first phoneme in the known correct alignment that corresponds to a second phoneme in the incorrect segment alignment;

modifying a first acoustic model of the first phoneme by moving at least one mean value thereof closer to feature values used to score the first phoneme.

13. (Amended) A method as recited in Claim 12, further comprising the steps of:

modifying a second acoustic model of the second phoneme by moving at least one mean value thereof farther from the feature values used to score the second phoneme.

14. (Amended) A computer-readable medium carrying one or more sequences of instructions for training acoustic models of a segmentation-based automatic speech recognition system, wherein execution of the one or more sequences of instructions by one or more processors causes the one or more processors to perform the steps of:

receiving correct alignment data that represents a correct segment alignment of an utterance that was received by the speech recognition system;

receiving wrong alignment data that represents an alignment of the utterance that is known to be incorrect based on information received from the speech recognition system and describing the utterance;

identifying a first phoneme in the correct alignment data that corresponds to a second phoneme in the wrong alignment data;

modifying a first acoustic model of the first phoneme by moving at least one mean value thereof closer to the feature values used to score the first phoneme.

15. (Amended) A computer-readable medium as recited in Claim 14, wherein the instructions further comprise instructions for carrying out the steps of:

modifying a second acoustic model of the second phoneme by moving at least one mean value thereof farther from the feature values used to score the second phoneme.

16. (Amended) A segmentation-based automatic speech recognition system that provides improved performance by training its acoustic models according to information about an application with which the system is used, comprising:

a recognizer that includes one or more processors;

non-volatile storage coupled to the recognizer and comprising a plurality of segmentation alignment data and a plurality of acoustic models;

a computer-readable medium coupled to the recognizer and carrying one or more sequences of instructions for the training acoustic models, wherein execution of the one or more sequences of instructions by the one or more processors causes the one or more processors to perform the steps of:

receiving correct alignment data that represents a correct segment alignment of an utterance that was received by the speech recognition system;

receiving wrong alignment data that represents an alignment of the utterance that is known to be incorrect based on information received from the speech recognition system and describing the utterance;

identifying a first phoneme in the correct alignment data that corresponds to a second phoneme in the wrong alignment data;

B2 modifying a first acoustic model of the first phoneme by moving at least one mean value thereof closer to the feature values used to score the first phoneme.

17. (Amended) A speech recognition system as recited in Claim 16, wherein the instructions further comprise instructions for carrying out the steps of:

modifying a second acoustic model of the second phoneme by moving at least one mean value thereof farther from the feature values used to score the second phoneme.

18. (New) A method of unsupervised training of acoustic models of a phonetic-based automatic speech recognition system, comprising the steps of:

receiving correct alignment data that represents a correct segment alignment of an utterance that was received by the speech recognition system;

receiving wrong alignment data that represents an alignment of the utterance that is known to be incorrect based on information received from the speech recognition system and describing the utterance;

A3 identifying a first phoneme in the correct alignment data that corresponds to a second phoneme in the wrong alignment data and in which the first phoneme received a worse recognizer score than the second phoneme;

modifying a first acoustic model of the first phoneme by moving at least one mean value thereof closer to the feature values used to score the first phoneme.

19. (New) A method as recited in Claim 18, further comprising the steps of:

modifying a second acoustic model of the second phoneme by moving at least one mean value thereof farther from the feature values used to score the second phoneme.

20. (New) A method as recited in Claim 18, wherein the first acoustic model includes a plurality of model components and wherein modifying a first acoustic model further comprises the steps of modifying a set of the model components associated with the first phoneme by moving all mean values thereof closer to the corresponding feature values used to score the phoneme.

A3

Remarks

Reconsideration of this application is respectfully requested. Claims 1-3, 6-17 are amended. Claims 18-20 are new.

Applicants wish to thank the examiner for the personal interview.

All pending claims were rejected as anticipated by U.S. Pat. No. 6,272,462 (Nguyen) and/or unpatenable in view of Nguyen and U.S. Pat. No. 5,027,406 (Roberts). Applicants reserve their right to backdate Nguyen, if necessary, but believe that the pending claims require no such argument because of the clear distinctions.

Nguyen et al. discloses a supervised adaptation method (see, e.g., title and col. 1, ll. 38-49 and col. 3, ll. 1-5) that requires the correct transcription for training to be known in advance. Moreover, once adaptation is triggered the entire sentence and/or words are adapted (see, e.g., col. 3, l. 50-56).

In contrast to Nguyen, the present claims specifically refer to an unsupervised adaptation method and system. Unsupervised adaptation means that the models are adapted without the benefit of a training script that is known a priori. This distinction is basic and clear over Nguyen which is specific to supervised adaptation. Moreover, the new claims make clear that adaptation discriminates and trains on a phoneme-basis. Thus, the set of phonemes which are most likely to benefit from training are identified and adapted resulting in enhanced effectiveness and efficiency. No such teaching or suggestion is provided by Nguyen, which operates on a coarser scale of adaptation.

The claims were amended to clarify these distinctions. For example, the original claims included limitations directed to phoneme-level discrimination but were considered by the applicants, after the benefit of the personal interview, as perhaps unclear. Thus, these limitations were clarified as were other limitations identifying "correct" versus "incorrect" and the like.

New claims 18-20 are added which are analogous to pending claims except they recite a phonetic based automatic speech system.

In view of the above amendments and comments, applicants believe that the claims are in a state proper for allowance and therefore urge the examiner to pass the claims to allowance. The examiner is encouraged to telephone the undersigned to discuss any matters in furtherance of the prosecution of the subject application.

Respectfully submitted,

Peter M. Dichiara
Reg. No. 38,005
Attorney for Applicant

Hale and Dorr LLP
60 State Street
Boston, MA 02109
(617) 526-6466
April 4, 2002

I hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail in an envelope addressed to: Commissioner of Patents and Trademarks, Washington, D.C. 20231

on 4-4-02
(Date of Deposit)

Signature

Attachment A

1. (Amended) A method of unsupervised training of acoustic models of a segmentation-based automatic speech recognition system, comprising the steps of:

receiving correct alignment data that represents a correct segment alignment of an utterance that was received by the speech recognition system;

receiving wrong alignment data that represents an alignment of the utterance that is known to be incorrect based on information received from the speech recognition system and describing the utterance;

identifying a first phoneme in the [wrong] correct alignment data that corresponds to a second phoneme in the [correct] wrong alignment data and in which the first phoneme received a worse recognizer score than the second phoneme;

modifying a first acoustic model of the first phoneme by moving at least one mean value thereof [further from] closer to the feature values used to score the first phoneme.

2. (Amended) A method as recited in Claim 1, further comprising the steps of:

[receiving correct alignment data that represents an alignment of the utterance that is known to be correct based on information received from the speech recognition system and describing the utterance;

identifying a second phoneme in the correct alignment data that corresponds to the first phoneme in the wrong alignment data;]

modifying a second acoustic model of the second phoneme by moving at least one mean value thereof [closer to] farther from the feature values used to score the second phoneme.

3. (Amended) A method as recited in Claim 1, wherein receiving correct alignment data comprises the step of receiving correct alignment data that represents a segment alignment of a less than highest scoring hypothesis [hypothesized alignment selected] from among n-best hypotheses of an utterance that was received by the speech recognition system.

4. (Unamended) A method as recited in Claim 1, wherein receiving wrong alignment data comprises the steps of receiving wrong alignment data that represents an alignment of the utterance that is known to be incorrect based on user confirmation information received from the speech recognition system in response to prompting a speaker to confirm the utterance.

5. (Unamended) A method as recited in Claim 1, wherein receiving correct alignment data comprises the steps of receiving correct alignment data that represents an alignment of the utterance that is known to be correct based on user confirmation information received from the speech recognition system in response to prompting a speaker to confirm the utterance.

6. (Amended) A method as recited in Claim 1, further comprising the step of iteratively repeating the identifying and modifying steps for all phonemes in the [wrong] correct alignment data that correspond to one or more phonemes in the [correct] wrong alignment data.

7. (Amended) A method as recited in Claim 2, further comprising the step of iteratively repeating the identifying and modifying steps for all phonemes in the [correct] wrong alignment data that correspond to one or more phonemes in the [wrong] correct alignment data.

8. (Amended) A method as recited in Claim 1, wherein the step of moving at least one mean value [further] farther from [a corresponding mean value of a second acoustic model of the second phoneme] the feature value used to score it comprises subtracting a multiple of the [mean] feature value [of the third acoustic model] from the mean value of the [second] first acoustic model.

9. (Amended) A method as recited in Claim 1, wherein the step of moving at least one mean value [further] farther from [a corresponding mean value of a second acoustic model of the second phoneme] the feature value used to score it comprises [reducing] modifying the mean value of the [third] first acoustic model by approximately two percent (2%).

10. (Amended) A method as recited in Claim 1, wherein the first acoustic model includes a plurality of model components and wherein modifying a first acoustic model further comprises the steps of modifying [all acoustic models] a set of the model components associated with the first phoneme by moving all mean values thereof [further from] closer to the corresponding [mean] feature values used to score the phoneme [of all second acoustic models associated with the second phoneme].

11. (Amended) A method as recited in Claim 2, wherein the second acoustic model includes a plurality of model components and wherein modifying [a third] the second acoustic model further comprises the steps of modifying [all acoustic models] a set of model components associated with the [third] second phoneme by moving all mean values thereof [closer to] farther from the corresponding [mean] feature values used to score the phoneme [of all acoustic models associated with the second phoneme]

12. (Amended) A method of improving performance of a segmentation-based automatic speech recognition system (ASR) by training its acoustic models using information obtained from a particular application in which the ASR is used, comprising the steps of:

receiving a correct segment alignment of an utterance that was received by the ASR;

receiving an incorrect alignment of the utterance that is known to be incorrect based on information received from the speech recognition system in the context of the particular application;

identifying a first phoneme in the known [incorrect] correct alignment that corresponds to a second phoneme in the [correct] incorrect segment alignment;

modifying a first acoustic model of the first phoneme by moving at least one mean value thereof [further from a corresponding mean value of a second acoustic model of the second phoneme] closer to feature values used to score the first phoneme.

13. (Amended) A method as recited in Claim 12, further comprising the steps of:

[receiving an alignment of the utterance that is known to be correct based on information received from the speech recognition system in the context of the particular application;

identifying a third phoneme in the known correct alignment that corresponds to the second phoneme in the correct alignment;]

modifying a [third] second acoustic model of the [third] second phoneme by moving at least one mean value thereof [closer to the corresponding mean value of the second acoustic model of the second phoneme] farther from the feature values used to score the second phoneme.

14. (Amended) A computer-readable medium carrying one or more sequences of instructions for training acoustic models of a segmentation-based automatic speech recognition system, wherein execution of the one or more sequences of instructions by one or more processors causes the one or more processors to perform the steps of:

receiving correct alignment data that represents a correct segment alignment of an utterance that was received by the speech recognition system;

receiving wrong alignment data that represents an alignment of the utterance that is known to be incorrect based on information received from the speech recognition system and describing the utterance;

identifying a first phoneme in the [wrong] correct alignment data that corresponds to a second phoneme in the [correct] wrong alignment data;

modifying a first acoustic model of the first phoneme by moving at least one mean value thereof [further from a corresponding mean value of a second acoustic model of the second phoneme] closer to the feature values used to score the first phoneme.

15. (Amended) A computer-readable medium as recited in Claim 14, wherein the instructions further comprise instructions for carrying out the steps of:

[receiving an alignment of the utterance that is known to be correct based on information received from the speech recognition system in the context of the particular application;

identifying a third phoneme in the known correct alignment that corresponds to the second phoneme in the correct alignment;]

modifying a [third] second acoustic model of the [third] second phoneme by moving at least one mean value thereof [closer to the corresponding mean value of the second acoustic model of the second phoneme] farther from the feature values used to score the second phoneme.

16. (Amended) A segmentation-based automatic speech recognition system that provides improved performance by training its acoustic models according to information about an application with which the system is used, comprising:

a recognizer that includes one or more processors;

non-volatile storage coupled to the recognizer and comprising a plurality of segmentation alignment data and a plurality of acoustic models;

a computer-readable medium coupled to the recognizer and carrying one or more sequences of instructions for the training acoustic models, wherein execution of the one or more sequences of instructions by the one or more processors causes the one or more processors to perform the steps of:

receiving correct alignment data that represents a correct segment alignment of an utterance that was received by the speech recognition system;

receiving wrong alignment data that represents an alignment of the utterance that is known to be incorrect based on information received from the speech recognition system and describing the utterance;

identifying a first phoneme in the [wrong] correct alignment data that corresponds to a second phoneme in the [correct] wrong alignment data;

modifying a first acoustic model of the first phoneme by moving at least one mean value thereof [further from a corresponding mean value of a second acoustic model of the second phoneme] closer to the feature values used to score the first phoneme.

17. (Amended) A speech recognition system as recited in Claim 16, wherein the instructions further comprise instructions for carrying out the steps of:

[receiving an alignment of the utterance that is known to be correct based on information received from the speech recognition system in the context of the particular application;

identifying a third phoneme in the known correct alignment that corresponds to the second phoneme in the correct alignment;]

modifying a [third] second acoustic model of the [third] second phoneme by moving at least one mean value thereof [closer to the corresponding mean value of the second acoustic model of the second phoneme] farther from the feature values used to score the second phoneme.

18. (New) A method of unsupervised training of acoustic models of a phonetic-based automatic speech recognition system, comprising the steps of:

receiving correct alignment data that represents a correct segment alignment of an utterance that was received by the speech recognition system;

receiving wrong alignment data that represents an alignment of the utterance that is known to be incorrect based on information received from the speech recognition system and describing the utterance;

identifying a first phoneme in the correct alignment data that corresponds to a second phoneme in the wrong alignment data and in which the first phoneme received a worse recognizer score than the second phoneme;

modifying a first acoustic model of the first phoneme by moving at least one mean value thereof closer to the feature values used to score the first phoneme.

19. (New) A method as recited in Claim 18, further comprising the steps of:

modifying a second acoustic model of the second phoneme by moving at least one mean value thereof farther from the feature values used to score the second phoneme.

20. (New) A method as recited in Claim 18, wherein the first acoustic model includes a plurality of model components and wherein modifying a first acoustic model further comprises the steps of modifying a set of the model components associated with the first phoneme by moving all mean values thereof closer to the corresponding feature values used to score the phoneme.